

## **A Method for Measuring Road Traffic Load Based on Analyzing Cellular Communications**

### **Field**

This invention relates generally to traffic load monitoring, and in particular to extracting speed and incident data out of the cellular phone system.

### **Background**

Kennedy (U.S. patent 5,657,487) teaches the use of handovers to determine vehicles speed and the number of vehicles driving on a certain route. Kennedy does not supply a solution to the very common problem in metropolitan areas of the same handovers relating to several different routes. This invention also discloses an extremely expensive implementation requiring RF receivers spread over the covered area.

Kaplan et al (WO 02/03350 A1) discloses a low cost, totally passive method (monitoring the cellular network without sending any messages) to extract traffic information from any cellular network in every load stage possible, with minimal hardware elements and hence minimal system cost.

It specifically teaches an initial differentiation and screening method to assign handover sequences to a specific route. This method will work perfectly when there is only one road within a covered area of several cells. However Kaplan et al does not give a full solution to the real life handover sequences received in urban areas. If a vehicle travels several times through the same route it will rarely generate the exact same handover sequence, furthermore the resemblance to handover sequences derived by traveling on different routes may be significant.

The current invention extends the work done by Kaplan et al and teaches detailed methods that efficiently and accurately correlate handover chains to a specific route, and avoid ambiguous correlations.

Furthermore the current invention teaches a method to determine the handover location accuracy for handovers within these chains.

These two steps generate a reference database that serves as a basis to locate the route on which vehicles are driving and their speed.

In addition the current invention discloses a method to correlate cellular calls by vehicles in the operational stage with the reference database and extract speed data and incident reports in real time.

## Summary of the Invention

The current invention describes a method for measuring traffic speed and travel time, as well as providing incident alerts, based on monitoring and analyzing cellular communications. Accurate mobile unit location is derived from the cellular communications data in specific positions and timings. This data is analyzed to derive the route on which the vehicle drives and its speed. Data is collected and analyzed for all routes continuously thus yielding real time traffic load and speed information as well as real time incident alerts.

## Brief Description of the Drawings

- Figure 1 shows the process of assigning unique chain clusters to specific routes and filtering ambiguous chains. Figures 1-A and 1-B show the cell sequences for streets 1 and 2 respectively. Using the following criteria: matching 3 out of 4 cells and requiring 3 trips at least in a cluster, 3 chain clusters were found for streets 1 and 2 as shown on figure 1-C. Chain clusters 1 and 3 are unique while chain cluster 2 is deleted by the ambiguity filter (3 out of 4 matches for different streets) therefore only 2 chain clusters exist in the unique chain clusters list shown in figure 1-D.
- Figure 2 demonstrates the handover accuracy definition. In figure 2-A the handover along the route (cell A to B) occurs as the building blocks the RF signal of cell A, the signal strength decays quickly, hence the handover from A to B will be accurate (has high location repeatability). In figure 2-B the handover along the route (C to D) occurs where there are no blocking elements so the signal from C decays slowly and the handover location error is much larger than the 1<sup>st</sup> handover. The handover location accuracy can be learned from the signal strength data or from the spread of the handover locations for different drives (marked by stars in the drawing).
- Figure 3 shows how handover locations are extracted from the operational drive data including only cell ID and time data. The learnt cell chains for street 1 are shown in figure 3-A. Figure 3-B shows the operational drive data. The initial chain matching shown in figure 3-C locates the route by using 3 out of 4 cell match. Once we have found the match we have certainty that we are located on street 1. This gives us handovers 2 and 3 on the handover table. Then we search for 2 out of 3 matches forward and backward and this search adds handovers 1 and 4 to the handover table as shown in figure 3-D. Since we know the location of each handover in the learnt chains and we have time for each handover from the operational chain as shown on figure 3-E we can extract the speed.
- Figure 4 shows how an incident that caused traffic to stop or slow down is detected without any additional handover data received. The learnt cell chains for street 2 are shown on figure 4-A. Operational drive data is shown on figure 4-B. The last handover was R to L.

Since then 1 minute have passed and we have not received additional cells and without receiving “end of call” message. Looking at figure 4-C we know that the longest distance to the next handover is 200 meters, so the maximal speed is 12 KM/H (shown on figure 4-D). When another minute passes without any new cell received the maximal speed goes down to 6 KM/H (figure 4-D).

- Figure 5-A describes the Handover flow rate over 3 route sections, A-B, B-C and C-D in regular, off peak hours, when traffic speed is around the maximal allowed speed. Figure 5-B describes the speed within these route sections as an incident occurs (T1-T5) and later on as it is cleared (T6-T10).
- Figure 6-A describes the affect of sector edge location accuracy and partial sector coverage on speed determination. Figure 6-B describes the speed determination over time.

## Description of the Invention

### Determination of Handover Locations and Signal Strength along Routes (Learning)

The 1<sup>st</sup> stage in the method is the learning stage, which is done prior to or in parallel with system installation as off-line processing. In this stage a database is created, containing handover (cell switch) sequences per route and can also contain location reference for each handover. This can be collected by actual drives on each route, and extracting the information from mobile units in the vehicles driving in any road and in any direction within a designated area.

The learning process through drives can be carried out in several ways, one of these ways is: vehicles with location determination system, such as synchronized GPS system, yielding continuously an accurate location and with operating cellular mobile unit travel across the designated area covering each road, in any driving direction and all possible turns. The handover information and the signal strength information for these mobile units are gathered and matched with the accurate location to yield the accurate location for each handover report or signal strength report.

This process may be repeated by driving several times throughout this area and by using several mobile units for each drive.

Each such drive creates for each mobile a sequence of handover reports accompanied with the accurate location of the vehicle per handover report. In addition each drive creates for each mobile a sequence of signal strength reports. Each such report includes the signal strength for several cells (up to 10) in addition to the accurate vehicle location.

All this raw data is kept in the reference database.

### Assigning Handover chains to specific routes

One of the ways to pre-process the recorded handover information, in order to filter ambiguous information, identify similarity criteria and save analysis time during operational stage, is as follows:

The handover raw data is processed and analyzed to create a database consisting of handover chain clusters that cover routes within the designated area.

A handover chain is a sequence of consecutive cell switches. In addition to the cell ID it may include for each handover (cell switch) the GPS location, the timing and any other relevant data. A chain cluster is a group of chains that is clustered according to a certain similarity rule. A sample of handover chains and chain clusters may be found in figure 1.

When we come to create this database we are faced with two phenomena:

- When driving several times, or with several mobile units, through the same route, the handover sequences generated in these drives are not identical.
- When driving through neighboring routes, in particular if they are parallel, similar (not

necessarily identical) handover chains may be generated.

These phenomena create two problems:

- Full matching of the same handover sequence in order to assign a route to a vehicle rarely occurs.
- We may erroneously assign a handover sequence to the wrong route.

One of the ways to overcome these problems is to perform the following processing:

All the drives through the same route section are processed to generate clusters of handover chains. These clusters include similar (not necessarily identical) handover chains. Ambiguous handover chains are filtered. Ambiguous chains are defined as similar (not necessarily identical) handover chains relating to different route sections. A fast indexing mechanism can be then built to enable fast access to the chain clusters database.

The objective of this processing is to create a cluster database consisting of all chains that conform to specific route sections. These clusters are unique, in the sense that there is no chain in the raw data that can be assigned to two different clusters representing two different route sections. This processing is performed once, within the learn process and lays the ground to fast, accurate and ambiguity free route assignment during the operational stage.

One of the preferred embodiments of this clustering and ambiguity filtration algorithm is as follows:

We cluster handover chains with  $L$  cells in a chain, All chains within a cluster are characterized in that they contain at least  $K$ ,  $K \leq L$ , cells that are identical to cells in a specific chain, designated as the cluster header. These  $K$  cells must be ordered in the same sequence within the chain and the cluster header. Each of these chains may include up to  $(L-K)$  cells that are not identical to cells in the header.

If a certain route was covered  $M$  times ( $M$  is the sum of the number of mobile units participating in the learning process per drive for all drives per route) We will define an acceptable cluster for a route section within the designated area as a cluster that includes at least  $N$  chains ( $N \leq M$ ,  $N > 1$ ).

An example to the chain clustering process can be found in figure 1. The raw data includes inputs from 4 ( $M$ ) drives in 2 streets (figures 1-A and 1-B). Figure 1-C shows the clusters created from the raw data which include at least 4 ( $N$ ) chains, that have a match of 3 ( $K$ ) out of 4 ( $L$ ) cells.

All chains within the clusters are then processed against the raw data, another embodiment of the invention teaches initial chains processing against other chains. If a chain is found in the raw data or the chain cluster that includes at least  $J$  ( $J \leq L$ ) cells identical to the cells in the processed chain ordered in the same sequence, and this chain relates to a different route section than the cluster being processed, then the cluster is marked as ambiguous and erased from the clusters database.

An example for this filtration process is shown in figures 1-C and 1-D whereas cluster no. 2 includes chains both from street 1 and street 2 and it is discarded and not included in the unique chain clusters.

Different routes within the designated area will accept different J,K,L,M and N values, according to the route topography and density, cell coverage etc.

In order to enable very fast access to the clusters database a fast access mechanism (such as hashing) can be implemented. The fast access system is built once during the learning stage and updated afterwards if necessary.

If the area analyzed by this algorithm is small enough and the processing power is large enough, this ambiguity analysis can be conducted in real time, in conjunction with a handovers sequence received during the operational stage, and provide correlation result such as detailed in Operational stage below.

#### Defining handover accuracy

If very accurate location and speed information is needed, further analysis is needed to measure and define the accuracy level for each handover in the chain clusters.

This can be done in several ways, such as:

- By using the signal strength pattern for the relevant cells
- By measuring the location spread of handovers between the same cells for different chains within the cluster (different trips).
- Any combination of the above two ways.

The accuracy level per handover can be kept within the cluster database.

One of the preferred embodiments of this accuracy defining algorithm is as follows:

Each handover is performed between a submitting cell and a receiving cell. We follow the signal strength of the submitting cell around the handover location. If the power of the submitting cell decays very quickly (typically as a result of RF blocking and reflecting elements) then this handover is assigned a high accuracy level. If the power of the submitting cell decays very slowly then this handover is assigned a low accuracy level.

If the phenomena in the 1<sup>st</sup> option occurs more than once within the route, it still may be used for location reference if the different appearances can be distinguished by different sequences in which they appeared in.

Another preferred embodiment of the accuracy defining algorithm is to evaluate the spread of similar handover locations. For example: find the average of all similar handovers within a cluster, and decide whether this is an accurate handover according to the maximal location distance of the handovers from the average

An example for different infrastructure and its effect on the handover location accuracy is given in figure 2. The handover location for all similar handovers within a cluster are processed to generate the actual distance between the handover locations, the average handover location and an indicator for handover accuracy. All this data is kept in the learned database.

#### Learning stage for highways only

In case of detecting speed and incident of highways only, and there are only few highways in the relevant sub area that differ in volume of traffic at least for some time, the learning stage can be done without driving mapping procedure at all. In this case the system records the sequences over the relevant area for a significant time period to enable statistical analysis. Traffic volume information is fed into the system from external sources such as government reports and bureau of statistics. The volume of correlated sequences and the volume of traffic is then analyzed to correlate between clusters of sequences and the highways route sections, and to filter sequences of small roads around it.

Other sources of information can be used instead of external traffic volume information, such as the location of the cellular antennas, information from a single drive on the highways etc.

In such cases the system can collect virtually all possible chains for every highway and insert them to the learnt database.

#### Operational stage - Correlating Handover sequences with routes and call follow up.

During the operational stage, events related to calls on the cellular system are monitored.

The monitoring can be performed from the base stations (BS), base station controllers (BSC), main switching systems (MSC) or communication links between them.

One of the preferred embodiments of the current invention is by connecting to the links between the BSC and the MSC near the MSC. This type of connection is highly efficient in the sense that it involves connection to a minimal number of links and all of them are on the same physical location. Since the relevant information that can be extracted from these links is limited to handovers and their timing, it is one of the major breakthroughs of this invention, that enables extracting traffic data from such limited information.

Another preferred embodiment of the current invention is by extracting the handover and call processing related messages only thus minimizing the scope of information to be handled.

In such case, if the traffic extraction system is loaded with too much processing, it can extract only part of the calls when the cellular system or parts of it (a certain location area or BSC) is loaded to minimize the processing, where as these parts of the calls are enough in these parts of the cellular system for accurate and full location, speed and traffic load detection.

These events can be extracted and handled in real time, and there is no need to wait for a call to terminate in order to handle its events but rather handle each event as soon as it is extracted. This has a very high importance in incident detection as detailed below.

The learnt database is used to match the operational calls. This data may reside on a hard disk or in the computer's memory for faster response times.

The monitored events include, among others, the following events:

- Call processing events (start, end, call waiting hold, resume etc.)
- Location related events (Handovers, location area change etc.)

A call number and mobile ID (that can't identify the mobile phone participating in the call in order to avoid privacy violation) can be attached to each call to enable following the call and the mobile as they proceed across the covered area.

The location related events are used to pinpoint the route section which the vehicle is driving on and to determine the vehicle location (and location accuracy) in certain locations during the call.

The initial rough location of a vehicle can be done using the location area data or the actual link that supplies the data.

In order to identify the route section of the vehicle with a very high probability of detection on one hand and perform the initial detection quickly and efficiently on the other hand, we can use mechanisms such as the Unique Chain Clusters for initial route identification. We can also access the clusters database using the fast access mechanism.

In order to have a high probability of detection we can take a chain of P cells generated in the operational stage with at least Q ( $Q \leq P$ ) cells matching those in the clusters database where the cells in both chains are ordered in the same order. Of course P and Q may change in values, according to the route topography and density, cell coverage etc.

Figure 3 demonstrates the initial detection stage. Figure 3-A details the raw data and figure 3-B shown the relevant cell clusters. Figure 3-C shows the operational drive data and figure 3-D shows the initial matching of 3 (Q) out of 4 (P).

This match may not be found for each chain in the operational data. If this match is not found even for a single sub chain of a call than this call is not matched, and will not be used, unless new cells for this call will generate at least one sub chain that can be matched.

In order to achieve higher level of confidence against ambiguities, an optional stage can be conducted, in which the raw database is scanned to find a matching chain (Q out of P cells) to the operational drive chain at another route section. If such a chain is found then the current operational chain is discarded.

Once we have initial matching we can continue in parallel in two ways:

- Try to match handovers (cells) before or after this sub chain to match more cells for this call. This is done by following the raw data chains in the database looking for R out of S matches, where  $R \leq S$  and S is typically smaller than P (since we have high certainty of the route and we just need to match more cells between the learnt database and the current call). This secondary matching process is performed both forward for cells following the matched chain and backwards for cells preceding the matched chain. Figure 3-E shows this secondary matching process whereas the cells S, T, J and S from the operational drive (figure 3-C) are matched to trip 3 within the raw data (figure 3-A) using 2 (R) out of 3 (S)



- match, after the initial stage detected a match for the chain BCDG.
- Continue and find additional chain matches to sub chains within the current call.

If a chain which is a part of the current call is matched with more than one route in the learnt database, this part of the call is declared ambiguous and will not be used to extract speed, unless another part of it will be uniquely matched at a later stage, and this information will enable going backwards and re-matching the previous part of the call.

The matching can also be measured in similarity score, such as between 1 to 100, and the analysis for matching or filtering due to ambiguity can be conducted according to a matching threshold, that takes into consideration similarity of a sequence to several routes.

This threshold can vary in different sub-areas and according to the amount of data available for each route or sub-area in any given moment.

The next step can be to extract from the learnt database the locations of the handovers along the matched route sections, along with their actual timing and location accuracy. This stage is demonstrated in figure 3-F.

The matching procedure between the learnt database and the new chain can be done without building clusters, but rather by comparing the new chain directly with the raw chains in the learnt database. This way the processing required during the operational stage is larger, and may cause some time delays in providing the traffic information.

Following this procedure for many calls within the covered area continuously results in a stream of times, route sections, locations on the route and location accuracies for each call.

It is worth to note that at this stage the data does not include any cellular system related information, So from now on a cellular operator will not be sensitive to use or dissemination of this data.

The system can thus be divided into 2 parts: the 1<sup>st</sup> part that performs all processing that requires cellular operator confidential data such as handovers and the 2<sup>nd</sup> part that handles only location and timing data, The 2<sup>nd</sup> part can use other sources of information, such as Information from vehicle monitoring systems, to monitor traffic status.

#### Operational stage - Incident Detection

Real time incident detection is crucial for incident management, as well as for traffic routing.

When an incident occurs vehicles might significantly slow down or completely stop, and the flow of handovers from these vehicles might stop.

One of the preferred embodiments of the current invention enables to detect incidents in real time. One of the ways to it can be as follows:

The next possible cells for each call are extracted from the database, and their distance and expected timing (according to the vehicle speed) from the preceding cell is calculated. Once N seconds have passed since the last handover for a call and the next handover did not occur (and the call still continues) this means the

vehicle have slowed down or stopped. For this vehicle we can easily calculate the maximal possible speed for this route section.

An example to incident detection using this method is shown at figure 4. As the time passes ( $N=60$  for the 1<sup>st</sup> line in figure 4-D and 120 for the 2<sup>nd</sup> line in figure 4-D) and no new handover is received the maximal speed decreases (from 12 Km/H to 6 Km/H).

If this happened just for several vehicles concurrently this is a high probability indication for an incident along this route section. Since we know where the relevant vehicles are driving according to the correlation procedure we can detect the location of the incident.

#### Using handover rate for speed and incident status extraction.

In another embodiment of the current invention traffic speed on a route section is extracted in the following way:

1. In a calibration phase the speed along the route section is measured by external means (e.g. Driving along this route section and recording the speed) and in parallel the handovers (Cells switch) rate (handover time density) for this route section is monitored. The data for all route sections is kept in a database.
2. In the operational stage the handover rate along route sections is monitored and by comparing it to the rate measured during the calibration phase the actual speed along the route section is calculated.

The calibration phase can also be conducted without traffic speed measurements, by recording traffic density over a period of time on a route section, filtering extreme measurement samples, and then normalizing it according to the max average speed that is customary on this route section.

This method is a little less accurate for detecting speed, but is accurate enough to detect speed changes in a case of an incident.

This method can be used for incident detection where incident on a specific route section is detected when the handover rate over this route section decreases rapidly and an incident clearance is detected when the handover flow increases rapidly.

Figure 5 demonstrates incident occurrence and clearance detection. The incident occurs at T2 on section B-C and causes significant slow down on this section while A-B slows down in a delay (T3-T5) and C-D is not affected. When the incident is cleared at T6 B-C is cleared first (until T9) and A-B afterwards (until T10), while a temporary traffic slow down is created on C-D (at T6-T9) as the vehicles from A-B and B-C start moving forward but is cleared quickly.

#### Operational stage - Measuring speed and travel time in route sections and handling traffic incidents

At this stage the location data and incident occurrence and clearance reports per each route section can be

analyzed in order to extract speed and travel time, alarm on incidents along this route section and notify on incidents clearance.

Notification on incident clearance can be done whenever the speed over a route section increases significantly or whenever an incident clearance report is received. Incident clearance may be notified independently or may be notified only when an incident alarm was made for the same route section.

In addition we can extract the speed along several route sections if data from the same call or the same mobile unit covers several route sections. This is done because, for the same location accuracy in the start and end locations, the speed accuracy is higher as the distance between the start and end locations is larger.

This is demonstrated by figure 6-A. If the coverage of a 1 Km. Route section (A-F) is combined from coverage of sub-sections (calls 1-3 covering sub sections A-C, B-E and D-F) then the aggregate error is 180 meters ( $80+60+40=180$ ) and creates an error of  $\pm 90$  meters along the 1 Km. ( $\pm 9\%$ ). If on the other hand the speed is measured based on call 4 that covers the full section A-F, even with a high edge location accuracy (100 meter) the resulting speed accuracy will be  $\pm 5\%$ .

There are several factors to be taken into account when determining the speed accuracy per route section:

- The number of reports for this section.
- The timing of the location reports relative to current time (How “old” are the reports)
- The coverage percentage and overlap of the speed sub sections within the route section.
- The location accuracy of the edge locations.

Figure 6-A demonstrates the issue of location accuracy as discussed above. It involves a simple method of summing the location accuracy figures. Additional methods such as RMS summation can be applied. It also demonstrates partial coverage: If the only information includes calls 1 and 3 then sub section C-D is not covered and when we calculate the speed on the section A-F as a weighed average of the existing sub sections (A-C and D-F) then this is less accurate than an average covering the full section.

Figure 6-B demonstrates the importance of differentiating between old and new reports. If all reports between 8:40 and 9:00 are averaged the result is 68 Km/Hour, whereas if we apply the simple method of rolling 10 minutes average are calculated then the relative speed for 8:40-8:50, 8:45-8:55 and 8:50-9:00 are 65, 67 and 72 Km/Hour, and we can see the speed change over time. Additional methods such as weighed averaging may be used.

Using these criteria the speed along each route section can be calculated every T seconds (T may change as a function of time, number of reports etc. In figure 6-B T is 300 seconds or 5 minutes)

#### Operational stage for highways only

In case that traffic monitoring need to be conducted for highways only as described in the relevant section in the learning phase, the learnt database can virtually contain all possible sequences, and correlation procedure for new sequences is easier since the system can use simpler options for correlation, such as  $K =$

N and  $P=Q$ ,

#### Operational stage – Virtual sensor

One of the implementations of the current invention is in traffic control systems. These systems traditionally consist of road sensors that are connected to a control center. This kind of system is inferior to the method taught in the current invention because it has a much higher cost, measures only the momentary speed at the sensor location and is late at incident detection. Nevertheless, it will be required to implement the methods taught in the current invention as an addition to existing system based on traditional sensors. This will require significant changes in the control center to support such a hybrid system. In order to minimize the efforts and cost required to install such a hybrid system, even at the loss of some of the advantages mentioned above, virtual sensors are defined that extract from the cellular network the speed at several specific locations along the routes in the covered area and emulate the communication protocol between the control center and the road sensors in software.

#### Enhancing routes coverage by Operational stage data

During the operational stage large quantities of handover chains are received by the system and matched to the learnt database. The database can be enhanced to supply better coverage in the following manner:

In the 1st stage the system takes new chains that were matched to a certain route section and assigns each of the matched handovers in the new chain the location of the handover from the learnt database that was matched to it.

The rest of the handovers in the new chain, that were not assigned a location yet, are assigned locations now according to the following approximation:

- ◆ Each of these non-matched handovers, which is between 2 matched handovers, is assigned a location according to the distance from each of the matched handovers near it across the route. This distance is calculated assuming steady speed between the 2 matched handovers, i.e.  $\text{distance} = (\text{distance between the 2 matched handovers}) * (\text{time between matched handover and the non-matched handover}) / (\text{time between the 2 matched handovers})$ .
- ◆ If more than 1 non-matched handover is located between 2 matched handovers the calculation is done for each of the non-matched handovers separately according to the calculation above.

This new matched chain is then added to the learnt database and the handovers that were not matched are marked as potentially inaccurate.

If the same handover is assigned the same location in several different chains its accuracy level can be updated according to the variance of the different results.

#### Adjusting the learnt database to changes in the cellular system.

Cellular systems are changing from time to time and portions of the learnt database can become inappropriate. One of the preferred embodiments of the current invention is a method that minimizes the need for re-conducting the learning procedure by detecting changes in the cellular system and updating the learnt database by using data gathered during the operational stage.

The system can identify a chain cluster or chains in the learnt database, that statistically were matched X times per day, are matched only Y times per day, and  $Y \ll X$ , or even does not appear at all. This cluster or chains appear at a certain location across a certain route and has chains preceding or following it.

In such a case the system can trace new sequences located after the preceding or before following chains, looking for a new cluster of chains, that were not common before, or did not exist at all before, and their current appearance rate is very similar to X.

This new cluster of sequences is then assigned the route of the cluster that appears significantly less or does not appear at all.

The system can then try several similarity criteria to match handovers from the old cluster to handovers in the new cluster, in order to assign handovers in the new cluster with accurate location. Examples for such similarity algorithms are described previously in this invention.

Once several handovers from the new cluster are assigned location, the system may use the method from the previous section in order to assign location to the rest of the handovers.